Scientific Software Engineering & Data Management
X-ray Science Division (XSD), Advanced Photon Source
Effort Plan for FY17 and FY18

The APS Scientific Software Engineering & Data Management group provides leadership and scientific software engineering expertise in the areas of data analysis, data management, high-performance computing, visualization, mobile applications, and workflow and orchestration applications in support of world-class photon sciences at the APS. The mission is realized by the application of the group's core competencies in partnership with APS scientists and through collaborations with Laboratory and DOE resources, and the broader scientific community. Effort is aligned with facility priorities and is used in a transparent, flexible, and well-documented manner.

SDM believes it can uniquely enable great science (and x-ray synchrotron techniques) by creating great scientific and data management software. This is a key component in the creation of beamlines of the future, and is critical to the success of the APS Upgrade. This can be achieved with a core software application portfolio in prioritized areas, well-formed collaborative teams of software engineers, computer scientists, algorithm developers, beamline staff, users, peer groups, and others, and cutting-edge skills.

Focus areas for SDM effort in the coming years are well defined. Development effort is allotted to projects critical to the realization of techniques enabled by the APS Upgrade and aligned with the *APS Scientific Computing Strategy*. SDM will focus its effort on the development of scientific software for coherence imaging, fluorescence microscopy, time-dependent correlations, and high-energy diffraction, as well as on software tools for data access and management, and the automation of (near) real-time experiment feedback.

Project Summaries

Project	Summary	FY17 SDM FTE	FY18 SDM FTE
Bragg Coherent Diffraction Imaging (CDI) Software	Implement and parallelize genetic algorithms and phase retrieval methods for the Bragg CDI technique.	0.5	0.5
Correlation Toolkit	Develop a real-time HPC-enabled set of tools for time-based correlation data analysis.	0.75	0.75
General-Purpose Reciprocal-Space Mapping (RSM) Tools	Continue development and deployment of high-performance RSM tools.	1.0	1.0
X-ray Fluorescence Mapping (XFM) Software	Develop HPC-enabled fitting library and tools for fast elemental mapping.	0.75	0.75
Software and Tools for Tomographic Reconstruction	Support, maintain, develop, and accelerate software and tools for advanced tomographic reconstructions.	0.75	0.75
Data Quality & Feedback Tools	Toolkit and framework to verify quality of collected data and provide feedback during and after acquisition.	0.5	0.5

Workflow & Data Management Tools	Continue application of analysis workflow, web portals, and data management and distribution tools at APS beamlines.	0.5	0.5
Real-time Feedback / Data Acquisition (RTFB/DAQ) System	Software framework and tools for the collection of data used for controls, statistics and diagnostics of technical systems for the MBA accelerator.	0.75*	0.75*
Component Database (CDB)	An electronic system for tracking and documenting accelerator and beamline components.	1.0*	1.0*
Visualization Tools	Application and/or development of advanced visualization tools for APS beamline data analysis and experiment feedback.	*0.0*	*1.0***
Coherent Surface Scattering Imaging (CSSI) Software	Implementation of high-performance CSSI and GISAXS software applications.	*0.0*	*1.0***

* Partial or full funding from APS-U project based on FY16 ERA.
** Requires additional staffing effort / hires in the SDM group.
*** Work listed in this document describes projects requiring considerable effort. This list does not elaborate all of the SDM group's work, and the sum of the effort assignments is lower than the total number of individuals in the SDM group.

Project Details

Bragg Coherence Diffraction Imaging (CDI) Software

Summary: Implement and parallelize genetic algorithms and phase retrieval methods for Bragg CDI techniques.

Team: Barbara Frosik (SDM), Ross Harder (MIC), et al.
FY17 SDM Effort: 0.5 FTE
FY18 SDM Effort: 0.5 FTE

A genetic algorithm approach to CDI phase retrieval will improve coherent imaging in two aspects. The first is to enable the recovery of highly reproducible images from a given data set. The second is to render previously impossible to image samples amenable to CDI, opening the door to a greater scientific impact for the method. The basic idea is to do the same phasing process with tens to thousands of random starting points. The diversity of results is then exploited to arrive at a highly reproducible image of the sample. Another aspect of genetic algorithm approaches is in the "fitness" criterion used to evaluate the population of results. This can be tuned to enable phase retrieval of datasets that have previously been impossible to produce images from. It is desired to implement and parallelize software for fast processing by non-expert beamline users.

Current processing time of a 100 MB sample using serial MATLAB code takes 60 minutes using limited parameters. Current data acquisition time for a 100 MB data set is 20 minutes, and will decrease after the completion of the APS Upgrade. Attaining a robust image of a sample in a computation time nearer the data acquisition time will allow nearer real-time feedback into the experimental parameters. The experimenter may begin to do guided, carefully executed experiments. Currently, the vast majority of Bragg CDI users will benefit from semi-real-time phase retrieval for their data. It will also open the instrument up to far less sophisticated CDI users. This technique will be critical to one or more APS Upgrade beamlines.

Correlation Toolkit

Summary: Real-time HPC-enabled set of tools for time-based correlation data analysis.

Team: Faisal Khan (SDM), Suresh Narayanan (TRR), Alec Sandy (TRR), et al.
FY17 SDM Effort: 0.75 FTE
FY18 SDM Effort: 0.75 FTE

Time-based correlations are an important analysis tool used to study the dynamic nature of complex materials. The recent development and application of higher-frequency detectors allows the investigation of faster dynamic processes enabling novel science in a wide range of areas resulting in the creation of greater amounts of image data that must be processed within the time it takes to collect the next data set in order to guide data collection. The increased brightness afforded by the APS Upgrade project will compound this data processing challenge by producing data with higher count rates.

The current-generation HPC-enabled correlation system has been in use since February 2013, primarily by the Small-Angle X-ray Photon Correlation Spectroscopy (SA-XPCS) beamline at 8-ID-I, where it has been used by every user group at that beamline for multi-tau, and, more recently, two-time correlation analysis. These and other correlation algorithms will be critical to the techniques used by many APS Upgrade first experiments. In order to serve a larger community of dynamics driven experiments, such as image correlations for materials exploration and high-speed imaging techniques, a real-time, general-purpose correlation toolkit implementing more advanced correlation methods using parallel computational and algorithmic techniques, and deployed on high-performance computing (HPC) resources is required.

General-Purpose Reciprocal-Space Mapping (RSM) Tools

Summary: Continue development and deployment of high-performance RSM tools.

Team: John Hammonds (SDM), Jonathan Tischler (SSM), et al.
FY17 SDM Effort: 1.0 FTE
FY18 SDM Effort: 1.0 FTE

This project aims to continue development of a general-purpose tool for reciprocal-space mapping at the APS. The tool allows users to examine a volume of data and select portions on which to apply transformations that convert detector pixel locations from diffractometer geometry to reciprocal-space units, and then map pixel data on to a 3D reciprocal-space grid. It can map data acquired using 4- and 6-circle diffractometers, and with scans taken over angles or energy, and can operate via a graphical user interface, or in batch processing mode. Data too big to fit entirely into memory at one time is processed in smaller chunks and reassembled to form the final output volume, allowing users to process arbitrarily large input datasets.

This tool has the potential to serve an even larger number of APS beamlines, and will be critical to a number of APS Upgrade beamlines and high-energy diffraction experiments. It is currently in regular use for scattering and diffraction experiments at the 33-BM and 33-ID beamlines, for micro-diffraction analysis at 34-ID, and for time-resolved diffraction work at 7-ID. Development is underway for WA-XPCS analysis at 8-ID, and for data exploration with inelastic x-ray measurements at 30-ID. Fast tools for reciprocal-space mapping using distributed computing resources are needed to make nearer real-time decisions regarding the next set of data that is collected.

Software and Tools for Tomographic Reconstruction

Summary: Support, maintain, develop, and accelerate software and tools for advanced tomographic reconstructions.

Team: Ke Yue (SDM), Francesco De Carlo (IMG), Doga Gursoy (CXS), et al.
FY17 SDM Effort: 0.75 FTE
FY18 SDM Effort: 0.75 FTE

X-ray Fluorescence (XRF) Microscopy Software

Summary: High-performance computing (HPC) enabled fitting library and tools for fast elemental mapping of x-ray fluorescence microscopy software.

Team: Arthur Glowacki (SDM), Stefan Vogt (MIC), at al.
FY17 SDM Effort: 0.75 FTE
FY18 SDM Effort: 0.75 FTE

XRF imaging typically involves the creation and analysis of 3D data sets, where at each scan position the full spectrum is recorded. This allows one to later process the data in a variety of different approaches, e.g., by spectral region-of-interest (ROI) summation with or without background subtraction, principal component analysis, or fitting. Additionally, it is possible to sum up the per pixel spectra over selected spatial ROIs so as to improve the photon statistics in such a spectrum.

The XRF microscopy technique is a staple technique that will be used by many APS Upgrade beamlines in combination with other x-ray acquisition modalities. The increase in intensity and smaller spot size due to benefits of the APS Upgrade will drastically increase data size and data rates for this technique. In order to facilitate real-time data analysis and fast feedback for experiment steering, HPC-enabled implementations of common elemental mapping algorithms and data I/O schemes that facilitate streaming data, and appropriate user interfaces are required.

Data Quality & Feedback Tools

Summary: Toolkit and framework to verify quality of collected data and provide feedback during and after acquisition.

Team: Barbara Frosik (SDM), John Hammonds (SDM), Francesco De Carlo (IMG), Doga Gursoy (CXS), et al.
FY17 SDM Effort: 0.5 FTE
FY18 SDM Effort: 0.5 FTE

Confirming the quality of data that has been and is being collected is critical to the efficient and effective use of experiment beam time at the APS. Without such quality checks experimenters are collecting "in the dark" and may discover issues only after beam time is over. As data volumes and rates increase due to the further deployment of higher frame-rate detectors and the increase in flux and coherence afforded by the APS Upgrade project, and due to the complexity of novel experiments enabled by the APS Upgrade project, the ability to check data quality in real-time in order to alert experimenters to potential data collection issues, and to be able to provide feedback and automatically adjust experiment parameters is necessary.

These tools and framework provide a mechanism for monitoring data and applying appropriate and user definable quality checks to experiment data. It can verify that EPICS PV values are within expected ranges, monitor files and compare them against predefined schemas, and apply mean, standard deviation, and delta mean statistical methods to data. Further work is required in order to catalog results for use as a log, and as input to machine learning algorithms for real-time, automated feedback.

Workflow & Data Management Tools

Summary: Continue application of analysis workflow, and data management and distribution tools at APS beamlines.

Team: Faisal Khan (SDM), Arthur Glowacki (SDM), Sinisa Veseli (SDM), et al.
FY17 SDM Effort: 0.5 FTE
FY18 SDM Effort: 0.5 FTE

As data rates and volumes increase due to a combination of advances in detector technologies, increased use of multi-modal acquisition techniques, and the planning benefits of the APS Upgrade project, current manual data workflow and management mechanism will not be sufficient. The APS has a need for tools and infrastructure that automate analysis pipelines, maintain and track data ownership, catalog metadata, provides data distribution endpoints and Software as a Service (SaaS) web interfaces for data analysis, etc.

The APS team will place great emphasis on leveraging best-in-class tools, rather than on developing new systems. For example, they will continue to work closely with the Globus Services team in order to not duplicate effort and best leverage DOE and ANL resources. Open source tools will be used in order to best meet the needs of the APS in an efficient and cost effective manner. Of particular importance for the next two years is the development of online portals for XRF and XPCS analysis.

Real-time Feedback / Data Acquisition (RTFB/DAQ) System

Summary: Software framework and tools for the collection of data used for controls, statistics and diagnostics of technical systems for the MBA accelerator.

Proposed Team: Sinisa Veseli (SDM), Ned Arnold (AES Controls), John Carwardine (APS-U), et al.
FY17 SDM Effort: 0.75 FTE*
FY18 SDM Effort: 0.75 FTE*

The real-time feedback / data acquisition (RTFB/DAQ) system is a software framework and associated tools that enable fast data collection for controls, statistics, and diagnostics associated with the state-of-the-art embedded controllers utilized by the APS-U project MBA-based accelerator design. The DAQ software interfaces with several technical subsystems to provide time-correlated and synchronously sampled data that can be used for commissioning, troubleshooting, performance monitoring, and early fault detection. The key features of the system include capability to acquire data from multiple subsystems at various sample rates, support for continuous data acquisition, and the ability to route data to any number of applications. Future work will focus on extending system functionality to provide access to BPM turn-by-turn data, as well as power supply monitoring.

Component Database (CDB)

Summary: An electronic system for tracking and documenting accelerator and beamline components.

Proposed Team: Dariusz Jarosz (SDM), Sinisa Veseli (SDM), Ned Arnold (AES Controls), John Carwardine (APS-U), et al.
FY17 SDM Effort: 1.0 FTE*
FY18 SDM Effort: 1.0 FTE*

The Advanced Photon Source Upgrade (APSU) project aims to replace the existing APS storage ring under an aggressive one-year schedule for removal, installation and testing. Management of the thousands of components to be installed in such a short time represents a significant challenge. The Component Database (CDB) is a tool for organizing and tracking components and designs used for the APS storage ring upgrade. It helps capture component documentation, provides a repository for inspection and measurement data (e.g. electronic travellers), and supports logging of component history through the component's life cycle. CDB is built around relational database, web portal, and REST web service technologies, and provides users with a number of options for accessing the system. In particular, it serves as a user portal for finding all known information about a specific component or a design. To that end, CDB provides links and interfaces to external systems commonly used at APS, such as various drawing and document management systems, procurement applications, etc. Future efforts will focus on extending the system capabilities to capture cable connections, enabling access via mobile applications, and providing custom views for different user groups.

Visualization Tools

Summary: Application and/or development of advanced visualization tools for APS beamline data analysis and experiment feedback.

Proposed Team: TBD SDM Member, et al.
FY17 SDM Effort: 0.0 FTE**
FY18 SDM Effort: 1.0 FTE**

Visualization is often critical to experiment data analysis. Visualization of data from tomographic imaging, micro-diffraction, and high-energy diffraction beamlines is already a challenge that will become more pressing in the near future. With the increase in data volumes being generated by higher frame-rate detectors, and as novel multi-modal techniques are enabled due to the benefits of the APS Upgrade project and planned as a part of the APS Upgrade's first experiments, e.g. x-ray fluorescence microscopy data coupled with coherent diffraction imaging, advanced visualization techniques will be needed in order to gain understanding and insight from this data, both as a part of post-acquisition processing and to allow user intervention during data collection. The application, augmentation, and/or development of capable data visualization tools, such as ParaView, on advanced computational resources are needed in order to cope with these large and complex data streams.

Coherent Surface Scattering Imaging (CSSI) Software

Summary: Implementation of high-performance CSSI and GISAXS software applications.

Proposed Team: TBD SDM Member, Zhang Jiang (TRR), Jin Wang (TRR), et al.
FY17 SDM Effort: 0.0 FTE**
FY18 SDM Effort: 1.0 FTE**

CSSI and the non-coherence dependent complementary technique, GISAXS, will be critical techniques used at a number of APS Upgrade beamlines. Algorithm and model development are currently underway. Proposed LDRDs will further required development in this area, however, software coding effort will be needed in order to fully realize the benefit of these developments once sufficient algorithmic advances are made.